

Corpus-Comparative Methods for Cross-Cultural Analyses

1. Introduction

- The emerging fields of digital humanities and cultural analytics have not seen widespread adoption within the context of religious studies.
- The theoretical lenses provided by religious studies offer important opportunities for computational methods.
- 1) Religious studies has long grappled with themes highly relevant for our increasingly digital worlds such as the adoption of beliefs, formation of identities, etc.
- 2) The cross-cultural perspectives taken in religious studies are not confined to the study of religion in a strict sense, but are relevant for the study of culture more broadly.
- Broad goals of this work:
- 1) Demonstrate the usefulness of emerging methods from the computational study of culture for a religious studies context.
- 2) Demonstrate the usefulness of religious studies' theoretical lenses for developing novel methods for the computational study of culture.
- 3) Demonstrate the usefulness of this transdisciplinary work for broader study of culture beyond religion.

4. Computational Comparative **Religion and Reddit**

- Discussion text collected from various religion-oriented English-language communities on Reddit (currently 20 communities).
- Discussion text will likely not be representative of broader religious traditions, but of specific online communities.
- Several communities have a textual history of over a decade.
- Useful testing ground to implement comparisons.
- Lessons learned in this arena will inform the application of these methods outside the context of religion.

Zachary K. Stine email: zkstine@ualr.edu | twitter: @zacharykstine

2. Cross-Cultural Comparisons

- Three dimensions for cultural comparison:
 - 1) <u>Thematic</u>
 - What discursive themes bridge or distinguish cultural corpora?
 - Methods based in part on Klingenstein et al. (2014)
 - 2) <u>Dynamic</u>
 - How conceptually explorative or exploitative are cultural corpora?
 - Methods based in part on Barron et al. (2018) and Murdock et al. (2017). See recent usage for political systems in Stine & Agarwal (2019).
 - 3) Structural
 - What similarities/differences exist between cultural corpora after accounting for lexical differences?
 - Methods are believed to be novel and currently under development.
- Within each dimension of comparison,
- topic models used to reduce dimensionality of a cultural corpus into broader patterns of discourse;
- information theory utilized to interrogate relationships among corpora within semantic space of topic models.

5. Challenges and Evaluation

- Cultural comparisons do not correspond to fixed "groundtruth" relationships, but are contingent on a particular lens dictated by a particular methodology.
- Therefore, evaluation requires nuance in constructing evidence-based arguments.
- Metrics used in comparisons are relative; e.g. the distance between A and B is meaningful relative to both of their distances to C (and ideally others).
- The comparative framework underlying the structural dimension is abstract and therefore may be operationalized in a variety of manners.
- Overall effective evaluation will require careful reasoning, constant checking of the data to confirm intuition, and the input of subject matter experts.

3. Structural Comparisons

Lexical vs grammatical differences between cultures:

- worldviews of a culture.

- its conception of suffering.
- "interpret" another?
- become as lexical terms are removed?

6. Relevant Work

Barron, A.T.J., Huang, J., Spang, R.L., DeDeo, S.: Individuals, institutions, and innovation in the debates of the French Revolution. PNAS. 115, 4607–4612 (2018). https://doi.org/10.1073/pnas.1717729115.

1022 (2003).

Deitrick, J.E.: Engaged Buddhist Ethics: Mistaking the Boat for the Shore. In: Queen, C.S., Prebish, C.S., Keown, D. (eds.) Action Dharma: New Studies in Engaged Buddhism. pp. 252–269. Psychology Press (2003).

Klingenstein S., Hitchcock T., DeDeo, S.: The civilizing process in London's Old Bailey. PNAS. 111, 9419–9424 (2014). https://doi.org/10.1073/pnas.1405984111.

Manovich, L.: The science of culture? Social computing, digital humanities and cultural analytics. CA. (2016). https://doi.org/10.22148/16.004.

Murdock, J., Allen, C., DeDeo, S.: Exploration and exploitation of Victorian science in Darwin's reading notebooks. Cognition. 159, 117–126 (2017). https://doi.org/10.1016/j.cognition.2016.11.012.

Lövheim, M., Campbell, H.A.: Considering critical methods and theoretical lenses in digital religion studies. New Media & Society. 19, 5–14 (2017). https://doi.org/10.1177/1461444816649911.

Prothero, S.: Henry Steel Olcott and "Protestant Buddhism." Journal of the American Academy of Religion. 63, 281–302 (1995).

Stine, Z.K., Agarwal N.: A Quantitative Portrait of Legislative Change in Ukraine In: Thomson R., Bisgin H., Dancy C., Hyder A. (eds) Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2019. Lecture Notes in Computer Science, vol 11549. Springer, Cham. https://doi.org/10.1007/978-3-030-21741-9_6.

Collaborators:

Nitin Agarwal [adviser] (University of Arkansas at Little Rock – Department of Information Science) James E. Deitrick (University of Central Arkansas – Department of Philosophy and Religion)





Lexical differences—specialized vocabulary exists among the cultures that distinguishes or unites them. More surface-level, aesthetic features.

Grammatical differences—cultural level underlying lexicon corresponding to deeper structural elements of a culture. Closer to the underlying

Motivating work from religious studies:

Prothero (1995)—explores the notion of "Protestant Buddhism" in which a Buddhist lexicon is set upon a fundamentally Protestant grammar.

Deitrick (2003)—argues that American Engaged Buddhism represents a Buddhist lexicon coupled with a liberal Christian grammar with respect to

Topic mapping approach: How predictably does one model

Modified corpus approach: How similar do the discourses

Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. Journal of Machine Learning Research 3, 993–